Introducing Features

In the previous lesson, we took a look at some examples of *tabular data*:

SKU	Make	Color	Quantity	Price
908721	Guess	Blue	789	45.33
456552	Tillys	Red	244	22.91
789921	A&F	Green	387	25.92
872266	Guess	Blue	154	17.56

> ***Note:*** *We have been referring to this as a data table, but you will also see data in this format called a* ***matrix****. The term matrix is commonly used in mathematics, and it refers to a rectangular array—which, just like a table, contains data arranged in rows and columns.*

Recall that the columns in a table can be referred to as **features**. In the above example, `color` and `quantity` are *features* of the products. In the last lesson, we mentioned briefly that **feature engineering** is an important part of data preparation. In this lesson, we'll look at this topic in more detail.

In many cases, the set of initial features in the data is not enough to produce high quality trained machine learning models. You can use **feature engineering** to derive new features based on the values of existing features. This process can be as simple as applying a mathematical function to a feature (such as adding 1 to all values in an existing feature) or it can be as complex as training a separate machine learning model to create values for new features.

Once you have the features, another important task is selecting the features that are most important or most relevant. This process is called **feature selection**.

Many machine learning algorithms cannot accommodate a large number of features, so it is often necessary to do **dimensionality reduction** to decrease the number of features.

QUIZ QUESTION

Can you match each of these terms with its description?

Submit to check your answer choices!

DESCRIPTION	TERM
The columns of a data table or matrix; also known as fields, or variables.	Features
The phenomenon in which an ML algorithm is not capable of coping with very large numbers of features.	The curse of dimensionality
The process through which we create new features.	Feature engineering
The process through which we choose which features will be used in the model training process.	Feature selection

SUBMIT

NEXT